1. What does one mean by the term "machine learning"?

Machine learning is a part of Data Science which provides statistical algorithm to be implemented in the data using coding languages like python and R to predict or recommend any real life problems.

2. Can you think of 4 distinct types of issues where it shines?

The 4 distinct types of issues it shines are as follows:

* Movie Recommendation
* Prediction of credit card default, diabetic person etc.
* Multi level clustering for bring insights like sales data
* Time require for delivery person

3. What is a labelled training set, and how does it work?

A labelled training data set is a dataset which has a target column, this type of dataset make ML algorithm to set up a target dataset so that the test data/ validation data can be predicted based on training the labelled data.

4. What are the two most important tasks that are supervised?

Classification and regression

5.Can you think of four examples of unsupervised tasks?

Four common unsupervised tasks included clustering, visualization, dimensionality reduction , and association rule learning.

6. State the machine learning model that would be best to make a robot walk through various unfamiliar terrains?

Reinforced Learning, where the robot can learn from response of the terrain to optimize itself.

7. Which algorithm will you use to divide your customers into different groups?

K-mean Clustering

8. Will you consider the problem of spam detection to be a supervised or unsupervised learning problem?

Spam Detection can be supervised learning. There are number of independent features which need to be trained for performing a ML algorithm and then a prediction can be made.

9. What is the concept of an online learning system?

Reinforced Learning system, this would come up with the desired outcome and punishing the undesired ones. In general, a reinforcement learning agent is able to perceive and interpret its environment, take actions and learn through trial and error.

10. What is out-of-core learning, and how does it differ from core learning?

Out-of-core learning refers to the machine learning algorithms working with data that cannot fit into a single machine's memory but can easily fit into some data storage, such as a local hard disk or web repository.

11. What kind of learning algorithm makes predictions using a similarity measure?

K Nearest Neighbour (KNN)

12. What's the difference between a model parameter and a hyperparameter in a learning algorithm?

Model parameters are the kind of parameters that are urgent for defining a model, whereas hyper parameters are those features which we define to fine tune our model. This parameters are manually defined to the model for better training and prediction. Generally hyper parameters are used after considering the model performance.

13. What are the criteria that model-based learning algorithms look for? What is the most popular method they use to achieve success? What method do they use to make predictions?

Model based algorithm look for a two set of data which is called train data and test data, each set will be split up into dependent and independent data (X.Y). The dependent feature is called target column or y, where as independent feature will be called as X. Depending on these the model will predict optimum values. This involves using algorithms like linear regression, logistic regression, random forest, etc. trees to create an underlying model from which predictions can be made for new data points.

Model based learning algorithm search for the optimal value of parameters in a model that will give the best results for the new instances. We often use a cost function or similar to determine what the parameter value has to be in order to minimize the function.

14. Can you name four of the most important Machine Learning challenges?

* Data overfitting and underfitting
* Data leakage
* Class imbalance
* Inclusion of irrelevant data

15. What happens if the model performs well on the training data but fails to generalize the results to new situations? Can you think of three different options?

The above mentioned situation is a case of over fitting model. This makes the situation of low bias and high variance. We can avoid over fitting by:

* Using different techniques like ridge and lasso
* Feature engineering
* Analyse the dataset

16. What exactly is a test set, and why would you need one?

A test data is a subset of raw data, this is a dataset similar to train data having lower dimensionality in shape. This is used to predict the outcome or target , so that we can find the accuracy score or the probability score of its correctness compared to the actual target data from test data.

17. What is a validation set's purpose?

 If the test set is locked away, but you still want to measure performance on unseen data as a way of selecting a good hypothesis, then divide the available data (without the test set) into a training set and a validation set. A validation set can also be said a set of examples used to tune the parameters of a classifier, for example to choose the number of hidden units in a neural network.

18.What precisely is the train-dev kit, when will you need it, how do you put it to use?

The train-dev kit is typically used when developing and evaluating machine learning models to prevent overfitting to the training data. Overfitting occurs when a model performs well on the training data but poorly on new, unseen data. By using a separate dev set, the model's performance can be evaluated on data that it has not been trained on, allowing for better estimation of its real-world performance.

To use the train-dev kit, the original dataset is split into three parts: the training set, the development set, and the test set. The test set is used for final evaluation of the model's performance after it has been trained and fine-tuned using the train-dev split.

19. What could go wrong if you use the test set to tune hyperparameters?

If we use the test set to tune hyperparameters, we risk overfitting to the test data, which can result in a model that performs poorly on new, unseen data.